README.md

Hieu Trinh

6/22/2020

Initial data for research

This is the course project for the Getting and Cleaning Data Coursera course.

The raw dataset can be download from the link bellow:

https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

Tidy Data Criterias

The goals here is to transform the raw datasets and select only subset of variables that relavent for future processing and still satifying the following conditions:

- Each variable forms a column.
- Each observation forms a row.
- Each type of observational unit forms a table.

How `run_analysis.R` script works?

Obtain the raw dataset

1. **Download the data**

```
library("dplyr")
fromUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
fileName <- "Dataset.zip"
dataDir <- "data"
filePath <- downloadF(fromUrl, fileName, directory = dataDir)
# NOTE:
# 1) required "dplyr" package for computing average of each column ind DF
# 2) downloadF is a function defined earlier in the run_analysis.R
```

2. **Unpack the zip file and get the list of files in unzip folder**

```
unzip(filePath, exdir = dataDir)
## Find the unzip dataset directory
```

3. **Get the list of files in unziped folder**

```
datasetDir <- list.dirs(dataDir, recursive = FALSE)
## list all files from the Dataset.zip
files <- list.files(datasetDir, recursive = TRUE)
```

```
 [1] "activity_labels.txt"                          "features_info.txt"
 [3] "features.txt"                                 "README.txt"
 [5] "test/Inertial Signals/body_acc_x_test.txt"    "test/Inertial Signals/body_acc_y_test.txt"
```

```
 [7] "test/Inertial Signals/body_acc_z_test.txt"    "test/Inertial Signals/body_gyro_x_test.txt"
 [9] "test/Inertial Signals/body_gyro_y_test.txt"   "test/Inertial Signals/body_gyro_z_test.txt"
[11] "test/Inertial Signals/total_acc_x_test.txt"   "test/Inertial Signals/total_acc_y_test.txt"
[13] "test/Inertial Signals/total_acc_z_test.txt"   "test/subject_test.txt"
[15] "test/X_test.txt"                              "test/y_test.txt"
[17] "train/Inertial Signals/body_acc_x_train.txt"  "train/Inertial Signals/body_acc_y_train.txt"
[19] "train/Inertial Signals/body_acc_z_train.txt"  "train/Inertial Signals/body_gyro_x_train.txt"
[21] "train/Inertial Signals/body_gyro_y_train.txt" "train/Inertial Signals/body_gyro_z_train.txt"
[23] "train/Inertial Signals/total_acc_x_train.txt" "train/Inertial Signals/total_acc_y_train.txt"
[25] "train/Inertial Signals/total_acc_z_train.txt" "train/subject_train.txt"
[27] "train/X_train.txt"
```

4. **Look at the `README.txt` for all relavent data targeted files**
   - test group
     - `test/subject_test.txt` : Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30
     - `test/y_test.txt` : Test labels
     - `test/X_test.txt` : Test set
   - train group
     - `train/subject_train.txt` : Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30
     - `train/y_train.txt` : Training labels
     - `train/X_train.txt` : Training set
   - other
     - `activity_labels.txt` : Links the class labels with their activity name
     - `features.txt` : List of all features

Loading targeted files

1. **Read files into respective data frame**

```
subjectTestDF <- read.table(file.path(datasetDir, "test/subject_test.txt"), header = FALSE)
activityTestDF <- read.table(file.path(datasetDir, "test/y_test.txt"), header = FALSE)
featuresTestDF <- read.table(file.path(datasetDir, "test/X_test.txt" ), header = FALSE)

subjectTrainDF <- read.table(file.path(datasetDir, "train/subject_train.txt"), header = FALSE)
activityTrainDF <- read.table(file.path(datasetDir, "train/y_train.txt"), header = FALSE)
featuresTrainDF <- read.table(file.path(datasetDir, "train/X_train.txt"), header = FALSE)

activityLabels <- read.table(file.path(datasetDir,  "activity_labels.txt"), header = FALSE)
featureNames <- read.table(file.path(datasetDir, "features.txt"), head=FALSE)
```

2. **Exam the datasets properties**

```
str(subjectTestDF)
```

```
'data.frame':   2947 obs. of  1 variable:
 $ V1: int  2 2 2 2 2 2 2 2 2 2 ...
```

```
str(activityTestDF)
```

```
'data.frame':   2947 obs. of  1 variable:
 $ V1: int  5 5 5 5 5 5 5 5 5 5 ...
```

```
str(featuresTestDF)
```

```
'data.frame':   2947 obs. of  561 variables:
 $ V1  : num  0.257 0.286 0.275 0.27 0.275 ...
 $ V2  : num  -0.0233 -0.0132 -0.0261 -0.0326 -0.0278 ...
```

```
    str(subjectTrainDF)
```

```
'data.frame':   7352 obs. of  1 variable:
 $ V1: int  1 1 1 1 1 1 1 1 1 1 ...
```

```
    str(activityTrainDF)
```

```
'data.frame':   7352 obs. of  1 variable:
 $ V1: int  5 5 5 5 5 5 5 5 5 5 ...
```

```
    str(featuresTrainDF)
```

```
'data.frame':   7352 obs. of  561 variables:
 $ V1  : num  0.289 0.278 0.28 0.279 0.277 ...
 $ V2  : num  -0.0203 -0.0164 -0.0195 -0.0262 -0.0166 ...
```

```
    str(activityLabels)
```

```
'data.frame':   6 obs. of  2 variables:
 $ V1: int  1 2 3 4 5 6
 $ V2: chr  "WALKING" "WALKING_UPSTAIRS" "WALKING_DOWNSTAIRS" "SITTING" ...
```

```
    str(featureNames)
```

```
'data.frame':   561 obs. of  2 variables:
 $ V1: int  1 2 3 4 5 6 7 8 9 10 ...
 $ V2: chr  "tBodyAcc-mean()-X" "tBodyAcc-mean()-Y" "tBodyAcc-mean()-Z" "tBodyAcc-std()-X" ...
```

3. **Participants in each group**

```
unique(subjectTestDF$V1)
```

```
NOTE: there are 9 out of 30 volunteers participated in the test group
[1]  2  4  9 10 12 13 18 20 24
```

```
unique(subjectTrainDF$V1)
```

```
NOTE: there are 21 out of 30 volunteers participated in the training group
[1]  1  3  5  6  7  8 11 14 15 16 17 19 21 22 23 25 26 27 28 29 30
```

Merge test and training datasets

1. **Merge datasets groups by row**

```
subjectDF <- rbind(subjectTestDF, subjectTrainDF)
activityDF <- rbind(activityTestDF, activityTrainDF)
featuresDF <- rbind(featuresTestDF, featuresTrainDF)
```

2. **Merge previous intermediate datasets by column**

```
datasetDF <- subjectDF %>% cbind(activityDF) %>% cbind(featuresDF)
```

3. **Factorize the activity column with enum for readabilty**

```
activityDF$V1 <- factor(activityDF$V1, levels = activityLabels$V1, labels = activityLabels$V2)
```

4. **Create column names for the datasetDF using `c("subject", "activit")` and `activityLabels`**

```
selectedVariableNames <-
    datasetDFNames[grep("mean\\(\\)|std\\(\\)|subject|activity", datasetDFNames)]
```

5. **Update column names of the datasetDF using datasetDFNames**

```
names(datasetDF) <- datasetDFNames
str(datasetDF)
```

```
'data.frame':   10299 obs. of  563 variables:
 $ subject               : int  2 2 2 2 2 2 2 2 2 2 ...
 $ activity              : Factor w/ 6 levels "WALKING","WALKING_UPSTAIRS",..: 5 5 5 5 5 5 5 5 5 5
 $ tBodyAcc-mean()-X     : num  0.257 0.286 0.275 0.27 0.275 ...
 $ tBodyAcc-mean()-Y     : num  -0.0233 -0.0132 -0.0261 -0.0326 -0.0278 ...
```

Extract measurements that pertaining to mean or std activities

```
selectedVariableNames <- select only column name with contain("std", "mean"), "subject", "activity"
tidyDataSet <- subset(datasetDF, select = selectedVariableNames)
rm(datasetDF)

str(tidyDataSet)
# NOTE: columns number reduced to 68 from 563
```

```
'data.frame':   10299 obs. of  68 variables:
 $ subject                   : int  2 2 2 2 2 2 2 2 2 2 ...
 $ activity                  : Factor w/ 6 levels "WALKING","WALKING_UPSTAIRS",..: 5 5 5 5 5 5 5 5 5 5
 $ tBodyAcc-mean()-X         : num  0.257 0.286 0.275 0.27 0.275 ...
 $ tBodyAcc-mean()-Y         : num  -0.0233 -0.0132 -0.0261 -0.0326 -0.0278 ...
 $ tBodyAcc-mean()-Z         : num  -0.0147 -0.1191 -0.1182 -0.1175 -0.1295 ...
 $ tBodyAcc-std()-X          : num  -0.938 -0.975 -0.994 -0.995 -0.994 ...
 $ tBodyAcc-std()-Y          : num  -0.92 -0.967 -0.97 -0.973 -0.967 ...
 $ tBodyAcc-std()-Z          : num  -0.668 -0.945 -0.963 -0.967 -0.978 ...
 $ tGravityAcc-mean()-X      : num  0.936 0.927 0.93 0.929 0.927 ...
 $ tGravityAcc-mean()-Y      : num  -0.283 -0.289 -0.288 -0.293 -0.303 ...
 $ tGravityAcc-mean()-Z      : num  0.115 0.153 0.146 0.143 0.138 ...
 $ tGravityAcc-std()-X       : num  -0.925 -0.989 -0.996 -0.993 -0.996 ...
 $ tGravityAcc-std()-Y       : num  -0.937 -0.984 -0.988 -0.97 -0.971 ...
 $ tGravityAcc-std()-Z       : num  -0.564 -0.965 -0.982 -0.992 -0.968 ...
 $ tBodyAccJerk-mean()-X     : num  0.072 0.0702 0.0694 0.0749 0.0784 ...
 $ tBodyAccJerk-mean()-Y     : num  0.04575 -0.01788 -0.00491 0.03227 0.02228 ...
 $ tBodyAccJerk-mean()-Z     : num  -0.10604 -0.00172 -0.01367 0.01214 0.00275 ...
 $ tBodyAccJerk-std()-X      : num  -0.907 -0.949 -0.991 -0.991 -0.992 ...
 $ tBodyAccJerk-std()-Y      : num  -0.938 -0.973 -0.971 -0.973 -0.979 ...
 $ tBodyAccJerk-std()-Z      : num  -0.936 -0.978 -0.973 -0.976 -0.987 ...
 $ tBodyGyro-mean()-X        : num  0.11998 -0.00155 -0.04821 -0.05664 -0.05999 ...
 $ tBodyGyro-mean()-Y        : num  -0.0918 -0.1873 -0.1663 -0.126 -0.0847 ...
 $ tBodyGyro-mean()-Z        : num  0.1896 0.1807 0.1542 0.1183 0.0787 ...
 $ tBodyGyro-std()-X         : num  -0.883 -0.926 -0.973 -0.968 -0.975 ...
 $ tBodyGyro-std()-Y         : num  -0.816 -0.93 -0.979 -0.975 -0.978 ...
 $ tBodyGyro-std()-Z         : num  -0.941 -0.968 -0.976 -0.963 -0.968 ...
 $ tBodyGyroJerk-mean()-X    : num  -0.2049 -0.1387 -0.0978 -0.1022 -0.0918 ...
 $ tBodyGyroJerk-mean()-Y    : num  -0.1745 -0.0258 -0.0342 -0.0447 -0.029 ...
 $ tBodyGyroJerk-mean()-Z    : num  -0.0934 -0.0714 -0.06 -0.0534 -0.0612 ...
 $ tBodyGyroJerk-std()-X     : num  -0.901 -0.962 -0.984 -0.984 -0.988 ...
 $ tBodyGyroJerk-std()-Y     : num  -0.911 -0.956 -0.988 -0.99 -0.992 ...
 $ tBodyGyroJerk-std()-Z     : num  -0.939 -0.981 -0.976 -0.981 -0.982 ...
 $ tBodyAccMag-mean()        : num  -0.867 -0.969 -0.976 -0.974 -0.976 ...
 $ tBodyAccMag-std()         : num  -0.705 -0.954 -0.979 -0.977 -0.977 ...
 $ tGravityAccMag-mean()     : num  -0.867 -0.969 -0.976 -0.974 -0.976 ...
 $ tGravityAccMag-std()      : num  -0.705 -0.954 -0.979 -0.977 -0.977 ...
 $ tBodyAccJerkMag-mean()    : num  -0.93 -0.974 -0.982 -0.983 -0.987 ...
 $ tBodyAccJerkMag-std()     : num  -0.896 -0.941 -0.971 -0.975 -0.989 ...
 $ tBodyGyroMag-mean()       : num  -0.796 -0.898 -0.939 -0.947 -0.957 ...
```

```
 $ tBodyGyroMag-std()         : num  -0.762 -0.911 -0.972 -0.97 -0.969 ...
 $ tBodyGyroJerkMag-mean()    : num  -0.925 -0.973 -0.987 -0.989 -0.99 ...
 $ tBodyGyroJerkMag-std()     : num  -0.894 -0.944 -0.984 -0.986 -0.99 ...
 $ fBodyAcc-mean()-X          : num  -0.919 -0.961 -0.992 -0.993 -0.992 ...
 $ fBodyAcc-mean()-Y          : num  -0.918 -0.964 -0.965 -0.968 -0.969 ...
 $ fBodyAcc-mean()-Z          : num  -0.789 -0.957 -0.967 -0.967 -0.98 ...
 $ fBodyAcc-std()-X           : num  -0.948 -0.984 -0.995 -0.996 -0.995 ...
 $ fBodyAcc-std()-Y           : num  -0.925 -0.97 -0.974 -0.977 -0.967 ...
 $ fBodyAcc-std()-Z           : num  -0.636 -0.942 -0.962 -0.969 -0.978 ...
 $ fBodyAccJerk-mean()-X      : num  -0.9 -0.944 -0.991 -0.991 -0.991 ...
 $ fBodyAccJerk-mean()-Y      : num  -0.937 -0.969 -0.973 -0.972 -0.98 ...
 $ fBodyAccJerk-mean()-Z      : num  -0.924 -0.973 -0.972 -0.97 -0.983 ...
 $ fBodyAccJerk-std()-X       : num  -0.924 -0.962 -0.992 -0.992 -0.994 ...
 $ fBodyAccJerk-std()-Y       : num  -0.943 -0.98 -0.971 -0.975 -0.979 ...
 $ fBodyAccJerk-std()-Z       : num  -0.948 -0.981 -0.972 -0.981 -0.989 ...
 $ fBodyGyro-mean()-X         : num  -0.824 -0.923 -0.973 -0.972 -0.976 ...
 $ fBodyGyro-mean()-Y         : num  -0.808 -0.926 -0.981 -0.981 -0.98 ...
 $ fBodyGyro-mean()-Z         : num  -0.918 -0.968 -0.972 -0.967 -0.969 ...
 $ fBodyGyro-std()-X          : num  -0.903 -0.927 -0.973 -0.967 -0.974 ...
 $ fBodyGyro-std()-Y          : num  -0.823 -0.932 -0.977 -0.972 -0.977 ...
 $ fBodyGyro-std()-Z          : num  -0.956 -0.97 -0.979 -0.965 -0.97 ...
 $ fBodyAccMag-mean()         : num  -0.791 -0.954 -0.976 -0.973 -0.978 ...
 $ fBodyAccMag-std()          : num  -0.711 -0.96 -0.984 -0.982 -0.979 ...
 $ fBodyBodyAccJerkMag-mean() : num  -0.895 -0.945 -0.971 -0.972 -0.987 ...
 $ fBodyBodyAccJerkMag-std()  : num  -0.896 -0.934 -0.97 -0.978 -0.99 ...
 $ fBodyBodyGyroMag-mean()    : num  -0.771 -0.924 -0.975 -0.976 -0.977 ...
 $ fBodyBodyGyroMag-std()     : num  -0.797 -0.917 -0.974 -0.971 -0.97 ...
 $ fBodyBodyGyroJerkMag-mean(): num  -0.89 -0.952 -0.986 -0.986 -0.99 ...
 $ fBodyBodyGyroJerkMag-std() : num  -0.907 -0.938 -0.983 -0.986 -0.991 ...
```

Clean up the variable names to make them more descriptive

```
tidyVariableNames <- names(tidyDataSet)
tidyVariableNames <- gsub("^t", "time", tidyVariableNames)
tidyVariableNames <- gsub("^f", "frequency", tidyVariableNames)
tidyVariableNames <- gsub("Acc", "Accelerometer", tidyVariableNames)
tidyVariableNames <- gsub("Gyro", "Gyroscope", tidyVariableNames)
tidyVariableNames <- gsub("Mag", "Magnitude", tidyVariableNames)
tidyVariableNames <- gsub("BodyBody", "Body", tidyVariableNames)
tidyVariableNames <- gsub("-mean", "Mean", tidyVariableNames)
tidyVariableNames <- gsub("-std", "Std", tidyVariableNames)
tidyVariableNames <- gsub("[-()]", "", tidyVariableNames)
names(tidyDataSet) <- tidyVariableNames
str(tidyDataSet)
```

```
'data.frame':   10299 obs. of  68 variables:
 $ subject                                 : int  2 2 2 2 2 2 2 2 2 2 ...
 $ activity                                : Factor w/ 6 levels "WALKING","WALKING_UPSTAIRS",..: 5 5
 $ timeBodyAccelerometerMeanX              : num  0.257 0.286 0.275 0.27 0.275 ...
 $ timeBodyAccelerometerMeanY              : num  -0.0233 -0.0132 -0.0261 -0.0326 -0.0278 ...
 $ timeBodyAccelerometerMeanZ              : num  -0.0147 -0.1191 -0.1182 -0.1175 -0.1295 ...
 $ timeBodyAccelerometerStdX               : num  -0.938 -0.975 -0.994 -0.995 -0.994 ...
 $ timeBodyAccelerometerStdY               : num  -0.92 -0.967 -0.97 -0.973 -0.967 ...
 $ timeBodyAccelerometerStdZ               : num  -0.668 -0.945 -0.963 -0.967 -0.978 ...
 $ timeGravityAccelerometerMeanX           : num  0.936 0.927 0.93 0.929 0.927 ...
```

```
$ timeGravityAccelerometerMeanY              : num  -0.283 -0.289 -0.288 -0.293 -0.303 ...
$ timeGravityAccelerometerMeanZ              : num  0.115 0.153 0.146 0.143 0.138 ...
$ timeGravityAccelerometerStdX               : num  -0.925 -0.989 -0.996 -0.993 -0.996 ...
$ timeGravityAccelerometerStdY               : num  -0.937 -0.984 -0.988 -0.97 -0.971 ...
$ timeGravityAccelerometerStdZ               : num  -0.564 -0.965 -0.982 -0.992 -0.968 ...
$ timeBodyAccelerometerJerkMeanX             : num  0.072 0.0702 0.0694 0.0749 0.0784 ...
$ timeBodyAccelerometerJerkMeanY             : num  0.04575 -0.01788 -0.00491 0.03227 0.02228 ...
$ timeBodyAccelerometerJerkMeanZ             : num  -0.10604 -0.00172 -0.01367 0.01214 0.00275 ...
$ timeBodyAccelerometerJerkStdX              : num  -0.907 -0.949 -0.991 -0.991 -0.992 ...
$ timeBodyAccelerometerJerkStdY              : num  -0.938 -0.973 -0.971 -0.973 -0.979 ...
$ timeBodyAccelerometerJerkStdZ              : num  -0.936 -0.978 -0.973 -0.976 -0.987 ...
$ timeBodyGyroscopeMeanX                     : num  0.11998 -0.00155 -0.04821 -0.05664 -0.05999 ...
$ timeBodyGyroscopeMeanY                     : num  -0.0918 -0.1873 -0.1663 -0.126 -0.0847 ...
$ timeBodyGyroscopeMeanZ                     : num  0.1896 0.1807 0.1542 0.1183 0.0787 ...
$ timeBodyGyroscopeStdX                      : num  -0.883 -0.926 -0.973 -0.968 -0.975 ...
$ timeBodyGyroscopeStdY                      : num  -0.816 -0.93 -0.979 -0.975 -0.978 ...
$ timeBodyGyroscopeStdZ                      : num  -0.941 -0.968 -0.976 -0.963 -0.968 ...
$ timeBodyGyroscopeJerkMeanX                 : num  -0.2049 -0.1387 -0.0978 -0.1022 -0.0918 ...
$ timeBodyGyroscopeJerkMeanY                 : num  -0.1745 -0.0258 -0.0342 -0.0447 -0.029 ...
$ timeBodyGyroscopeJerkMeanZ                 : num  -0.0934 -0.0714 -0.06 -0.0534 -0.0612 ...
$ timeBodyGyroscopeJerkStdX                  : num  -0.901 -0.962 -0.984 -0.984 -0.988 ...
$ timeBodyGyroscopeJerkStdY                  : num  -0.911 -0.956 -0.988 -0.99 -0.992 ...
$ timeBodyGyroscopeJerkStdZ                  : num  -0.939 -0.981 -0.976 -0.981 -0.982 ...
$ timeBodyAccelerometerMagnitudeMean         : num  -0.867 -0.969 -0.976 -0.974 -0.976 ...
$ timeBodyAccelerometerMagnitudeStd          : num  -0.705 -0.954 -0.979 -0.977 -0.977 ...
$ timeGravityAccelerometerMagnitudeMean      : num  -0.867 -0.969 -0.976 -0.974 -0.976 ...
$ timeGravityAccelerometerMagnitudeStd       : num  -0.705 -0.954 -0.979 -0.977 -0.977 ...
$ timeBodyAccelerometerJerkMagnitudeMean     : num  -0.93 -0.974 -0.982 -0.983 -0.987 ...
$ timeBodyAccelerometerJerkMagnitudeStd      : num  -0.896 -0.941 -0.971 -0.975 -0.989 ...
$ timeBodyGyroscopeMagnitudeMean             : num  -0.796 -0.898 -0.939 -0.947 -0.957 ...
$ timeBodyGyroscopeMagnitudeStd              : num  -0.762 -0.911 -0.972 -0.97 -0.969 ...
$ timeBodyGyroscopeJerkMagnitudeMean         : num  -0.925 -0.973 -0.987 -0.989 -0.99 ...
$ timeBodyGyroscopeJerkMagnitudeStd          : num  -0.894 -0.944 -0.984 -0.986 -0.99 ...
$ frequencyBodyAccelerometerMeanX            : num  -0.919 -0.961 -0.992 -0.993 -0.992 ...
$ frequencyBodyAccelerometerMeanY            : num  -0.918 -0.964 -0.965 -0.968 -0.969 ...
$ frequencyBodyAccelerometerMeanZ            : num  -0.789 -0.957 -0.967 -0.967 -0.98 ...
$ frequencyBodyAccelerometerStdX             : num  -0.948 -0.984 -0.995 -0.996 -0.995 ...
$ frequencyBodyAccelerometerStdY             : num  -0.925 -0.97 -0.974 -0.977 -0.967 ...
$ frequencyBodyAccelerometerStdZ             : num  -0.636 -0.942 -0.962 -0.969 -0.978 ...
$ frequencyBodyAccelerometerJerkMeanX        : num  -0.9 -0.944 -0.991 -0.991 -0.991 ...
$ frequencyBodyAccelerometerJerkMeanY        : num  -0.937 -0.969 -0.973 -0.972 -0.98 ...
$ frequencyBodyAccelerometerJerkMeanZ        : num  -0.924 -0.973 -0.972 -0.97 -0.983 ...
$ frequencyBodyAccelerometerJerkStdX         : num  -0.924 -0.962 -0.992 -0.992 -0.994 ...
$ frequencyBodyAccelerometerJerkStdY         : num  -0.943 -0.98 -0.971 -0.975 -0.979 ...
$ frequencyBodyAccelerometerJerkStdZ         : num  -0.948 -0.981 -0.972 -0.981 -0.989 ...
$ frequencyBodyGyroscopeMeanX                : num  -0.824 -0.923 -0.973 -0.972 -0.976 ...
$ frequencyBodyGyroscopeMeanY                : num  -0.808 -0.926 -0.981 -0.981 -0.98 ...
$ frequencyBodyGyroscopeMeanZ                : num  -0.918 -0.968 -0.972 -0.967 -0.969 ...
$ frequencyBodyGyroscopeStdX                 : num  -0.903 -0.927 -0.973 -0.967 -0.974 ...
$ frequencyBodyGyroscopeStdY                 : num  -0.823 -0.932 -0.977 -0.972 -0.977 ...
$ frequencyBodyGyroscopeStdZ                 : num  -0.956 -0.97 -0.979 -0.965 -0.97 ...
$ frequencyBodyAccelerometerMagnitudeMean    : num  -0.791 -0.954 -0.976 -0.973 -0.978 ...
$ frequencyBodyAccelerometerMagnitudeStd     : num  -0.711 -0.96 -0.984 -0.982 -0.979 ...
$ frequencyBodyAccelerometerJerkMagnitudeMean: num  -0.895 -0.945 -0.971 -0.972 -0.987 ...
```

```
 $ frequencyBodyAccelerometerJerkMagnitudeStd : num  -0.896 -0.934 -0.97 -0.978 -0.99 ...
 $ frequencyBodyGyroscopeMagnitudeMean        : num  -0.771 -0.924 -0.975 -0.976 -0.977 ...
 $ frequencyBodyGyroscopeMagnitudeStd         : num  -0.797 -0.917 -0.974 -0.971 -0.97 ...
 $ frequencyBodyGyroscopeJerkMagnitudeMean    : num  -0.89 -0.952 -0.986 -0.986 -0.99 ...
 $ frequencyBodyGyroscopeJerkMagnitudeStd     : num  -0.907 -0.938 -0.983 -0.986 -0.991 ...
```

Creates a second, independent tidy data and write to file

From the tidyDataSet above, creates a second, independent tidy data set with the average of each variable for each activity and each subject.

```
tidyDataSet2 <- tidyDataSet %>%
    group_by(subject, activity) %>%
    dplyr::summarise(across(.cols = everything(), mean), .groups = "keep")

write.table(tidyDataSet2, "tidy_data.txt", row.names = FALSE, quote = FALSE)
```

2. **Look at the sructure of the second tidy data set tidyDataSet2**

```
head(tidyDataSet2)
```

```
# A tibble: 6 x 68
# Groups:   subject, activity [6]
  subject activity timeBodyAcceler... timeBodyAcceler... timeBodyAcceler... timeBodyAcceler... timeBody
    <int> <fct>               <dbl>            <dbl>            <dbl>            <dbl>            <dbl>
1       1 WALKING             0.277         -0.0174           -0.111           -0.284           0.114
2       1 WALKING...            0.255         -0.0240           -0.0973          -0.355          -0.002
3       1 WALKING...            0.289         -0.00992          -0.108            0.0300         -0.031
4       1 SITTING             0.261         -0.00131          -0.105           -0.977          -0.923
5       1 STANDING            0.279         -0.0161           -0.111           -0.996          -0.973
6       1 LAYING              0.222         -0.0405           -0.113           -0.928          -0.837
```

We can verify that the content have been saved to "tidy_dataset.txt" file.

Produce HTML and PDF output files

```
library("rmarkdown")
rmarkdown::render("README.md",output_format = "all")
```